MACHINE TRANSLATION DEVICE AND PROGRAM

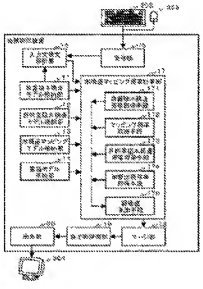
Publication number: JP2007018462
Publication date: 2007-01-25

Inventor: IMAMURA KENJI; OKUMA HIDEO; SUMIDA EIICHIRO
Applicant: ATR ADVANCED TELECOMM RES INST

Classification:
- international: ***G06F17/28;*** **G06F17/28;**
- European:
Application number: JP20050202350 20050711
Priority number(s): JP20050202350 20050711

Abstract of **JP2007018462**

PROBLEM TO BE SOLVED: To solve a problem of a conventional machine translation device incapable of performing high-quality translation at a high speed.
SOLUTION: Based on an original language tree structure model having original language tree structure information using a long unit phrase, which consists of a plurality of words and includes no non-terminal symbol, as one node, a target language tree structure model having target language tree structure information using a long unit phrase, which consists of a plurality of words and include no non-terminal symbol, as one node, a tree structure mapping model having mapping information, which shows correspondence between the original language structure and the target language tree structure information, and a mapping probability, and a language model having one or more word appearance probabilities serving as information about a probability concerned with appearance of a word in a second language, syntax analysis is carried out on a received sentence for acquiring a translation model probability and a language model probability. Based on these two probabilities, a syntax tree of an output is evaluated, and a translation sentence to be outputted is decided. In this way, this machine translation device can perform high-quality translation at a high speed.
COPYRIGHT: (C)2007,JPO&INPIT



(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2007-18462
(P2007-18462A)
(43)公開日　平成19年1月25日(2007.1.25)

(51)Int.Cl.		F I		テーマコード（参考）
G06F 17/28	(2006.01)	G06F 17/28	Q	5B091

審査請求 未請求　請求項の数 6　OL　（全 23 頁）

(21)出願番号　特願2005-202350 (P2005-202350)
(22)出願日　平成17年7月11日 (2005.7.11)

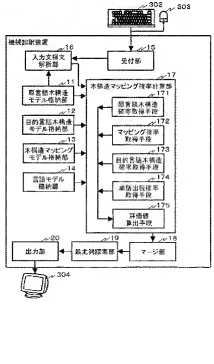
特許法第３０条第１項適用申請有り　２００５年３月１５日　言語処理学会発行の「言語処理学会　第１１回年次大会　発表論文集」に発表

(71)出願人　393031586
株式会社国際電気通信基礎技術研究所
京都府相楽郡精華町光台二丁目２番地２
(74)代理人　100115749
弁理士　谷川　英和
(72)発明者　今村　賢治
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内
(72)発明者　大熊　英男
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内
(72)発明者　隅田　英一郎
京都府相楽郡精華町光台二丁目２番地２
株式会社国際電気通信基礎技術研究所内

Fターム(参考)　5B091 AA04 BA14 CA05 CA21 CA24 EA02

(54)【発明の名称】機械翻訳装置、およびプログラム

(57)【要約】
【課題】従来の機械翻訳装置においては、高品質かつ高速な翻訳ができない、という課題があった。
【解決手段】複数の単語からなり非終端記号を含まない長単位句を一ノードとする原言語木構造情報を有する原言語木構造モデルと、複数の単語からなり非終端記号を含まない長単位句を一ノードとする目的言語木構造情報を有する目的言語木構造モデルと、原言語木構造情報と目的言語木構造情報との対応を示す情報であるマッピング情報とマッピング確率を有する木構造マッピングモデルと、第二言語における単語の出現に関する確率の情報である単語出現確率を１以上有する言語モデルとに基づいて、受け付けた文章を構文解析し、翻訳モデル確率と、言語モデル確率を取得し、当該２つの確率に基づいて、出力の構文木を評価し、出力する翻訳文を決定する機械翻訳装置により、高品質かつ高速な翻訳が可能となる。
【選択図】図１



【特許請求の範囲】
【請求項1】
翻訳される元の文章の言語である第一言語の木構造に関する情報であり、複数の単語からなり非終端記号を含まない長単位句を一のノードとする木構造の情報を含む原言語木構造情報と、当該原言語木構造情報に対応する木構造の確率を示す情報である原言語木構造確率を有する原言語木構造レコードを１以上有する原言語木構造モデルを格納している原言語木構造モデル格納部と、
翻訳結果の文章の言語である第二言語の木構造に関する情報であり、複数の単語からなり非終端記号を含まない長単位句を一のノードとする木構造の情報を含む目的言語木構造情報と、当該目的言語木構造情報に対応する木構造の確率を示す情報である目的言語木構造確率を有する目的言語木構造レコードを１以上有する目的言語木構造モデルを格納している目的言語木構造モデル格納部と、
原言語木構造情報と目的言語木構造情報との対応を示す情報であるマッピング情報と、当該マッピング情報が示す対応の確率を示す情報であるマッピング確率を有する木構造マッピングレコードを１以上有する木構造マッピングモデルを格納している木構造マッピングモデル格納部と、
第二言語における単語の出現に関する確率の情報である単語出現確率を１以上有する言語モデルを格納している言語モデル格納部と、
第一言語の文章を受け付ける受付部と、
前記受付部が受け付けた文章を構文解析し、当該文章の一部または全部の木構造に関する情報である木構造情報を、順次得る入力文構文解析部と、
前記入力文構文解析部が得た木構造情報に対応する１以上の原言語木構造確率を前記原言語木構造モデル格納部から取得し、前記入力文構文解析部が得た木構造情報に対応する１以上の目的言語木構造情報と１以上のマッピング確率を前記木構造マッピングモデル格納部から取得し、前記１以上の目的言語木構造情報のそれぞれに対応する１以上の目的言語木構造確率を前記目的言語木構造モデル格納部から取得し、前記取得した目的言語木構造情報に基づいて、当該目的言語木構造情報を構成する２以上の単語の、１以上の単語出現確率を前記言語モデル格納部から取得し、前記取得した原言語木構造確率、前記取得したマッピング確率、前記取得した目的言語木構造確率、および前記取得した単語出現確率に基づいて、出力の構文木の評価値を算出する木構造マッピング確率計算部と、
前記木構造マッピング確率計算部が算出した評価値に基づいて、出力する第二言語の一部または全部の構文木を決定し、当該決定した構文木に基づいて第二言語の一部または全部の文章である出力情報を取得する最尤列探索部と、
前記最尤列探索部が取得した１以上の出力情報を有する第二言語の文章を出力する出力部を具備する機械翻訳装置。
【請求項2】
前記木構造マッピング確率計算部は、
前記入力文構文解析部が得た木構造情報に対応する１以上の原言語木構造確率を前記原言語木構造モデル格納部から取得する原言語木構造確率取得手段と、
前記入力文構文解析部が得た木構造情報に対応する１以上の目的言語木構造情報と１以上のマッピング確率を前記木構造マッピングモデル格納部から取得するマッピング確率取得手段と、
前記１以上の目的言語木構造情報のそれぞれに対応する１以上の目的言語木構造確率を前記目的言語木構造モデル格納部から取得する目的言語木構造確率取得手段と、
前記取得した目的言語木構造情報に基づいて、当該目的言語木構造情報を構成する２以上の単語の、１以上の単語出現確率を前記言語モデル格納部から取得する単語出現確率取得手段と、
前記原言語木構造確率、前記マッピング確率、前記目的言語木構造確率、および前記単語出現確率に基づいて、出力の構文木の評価値を算出する評価値算出手段を具備する請求項

１記載の機械翻訳装置。
【請求項３】
前記木構造マッピング確率計算部が算出した２以上の評価値と当該評価値に対応する出力の構文木において、同一の構文木に対応する評価値を合成するマージ部をさらに具備し、
前記最尤列探索部は、
前記木構造マッピング確率計算部が算出した評価値および前記マージ部が合成して得た評価値に基づいて、出力する第二言語の一部または全部の構文木を決定し、当該決定した構文木に基づいて第二言語の一部または全部の文章である出力情報を取得する請求項１または請求項２記載の機械翻訳装置。
【請求項４】
前記評価値算出手段は、
前記原言語木構造確率、前記マッピング確率、前記目的言語木構造確率、前記単語出現確率の積を算出し、当該積を出力の構文木の評価値とする請求項２または請求項３記載の機械翻訳装置。
【請求項５】
前記最尤列探索部は、
前記木構造マッピング確率計算部が算出した評価値が最大の構文木を、出力する構文木として決定し、当該決定した構文木に基づいて第二言語の一部または全部の文章である出力情報を取得する請求項１から請求項４いずれか記載の機械翻訳装置。
【請求項６】
コンピュータに、
翻訳される元の文章の言語である第一言語の木構造に関する情報であり、複数の単語からなり非終端記号を含まない長単位句を一のノードとする木構造の情報を含む原言語木構造情報と、当該原言語木構造情報に対応する木構造の確率を示す情報である原言語木構造確率を有する原言語木構造レコードを１以上有する原言語木構造モデルを格納しており、
翻訳結果の文章の言語である第二言語の木構造に関する情報であり、複数の単語からなり非終端記号を含まない長単位句を一のノードとする木構造の情報を含む目的言語木構造情報と、当該目的言語木構造情報に対応する木構造の確率を示す情報である目的言語木構造確率を有する目的言語木構造レコードを１以上有する目的言語木構造モデルを格納しており、
原言語木構造情報と目的言語木構造情報との対応を示す情報であるマッピング情報と、当該マッピング情報が示す対応の確率を示す情報であるマッピング確率を有する木構造マッピングレコードを１以上有する木構造マッピングモデルを格納しており、
第二言語の２以上の単語の連続した出現に関する確率の情報である単語出現確率を有する単語出現確率を１以上有する言語モデルを格納しおり、
第一言語の文章を受け付ける受付ステップと、
前記受付ステップで受け付けた文章を構文解析し、当該文章の一部または全部の木構造に関する情報である木構造情報を、順次得る入力文構文解析ステップと、
前記入力文構文解析ステップで得た木構造情報に対応する１以上の原言語木構造確率を取得し、前記入力文構文解析ステップで得た木構造情報に対応する１以上の目的言語木構造情報と１以上のマッピング確率を取得し、前記１以上の目的言語木構造情報のそれぞれに対応する１以上の目的言語木構造確率を取得し、前記取得した目的言語木構造情報に基づいて、当該目的言語木構造情報を構成する２以上の単語の、１以上の単語出現確率を取得し、前記取得した原言語木構造確率、前記取得したマッピング確率、前記取得した目的言語木構造確率、および前記取得した単語出現確率に基づいて、出力の構文木の評価値を算出する木構造マッピング確率計算ステップと、
前記木構造マッピング確率計算ステップで算出した評価値に基づいて、出力する第二言語の一部または全部の構文木を決定し、当該決定した構文木に基づいて第二言語の一部または全部の文章である出力情報を取得する最尤列探索ステップと、

前記最尤列探索ステップで取得した１以上の出力情報を有する第二言語の文章を出力する、【発明の詳細な説明】せるためのプログラム。

【技術分野】

【0001】

本発明は、受け付けた文章を他言語に翻訳する機械翻訳装置等に関するものである。

【背景技術】

【0002】

従来の第一の機械翻訳装置において、「Ｐｈｒａｓｅ－ｂａｓｅｄ　ＳＭＴ」と呼ばれる翻訳アルゴリズムを採っていた（例えば、非特許文献１参照）。「Ｐｈｒａｓｅ－ｂａｓｅｄ　ＳＭＴ」とは、句に基づく統計翻訳であり、複数単語(句)を単位に翻訳を行う。

【0003】

従来の第二の機械翻訳装置において、「構文トランスファ方式ＭＴ」と呼ばれる翻訳アルゴリズムを採っていた（例えば、非特許文献２参照）。構文トランスファ方式の機械翻訳では、入力文を構文解析し、得られた構文木を出力の構文木に変換することにより翻訳を行う。

【0004】

さらに、従来の第三の技術として、本機械翻訳装置で利用され得る翻訳モデルの自動取得の技術がある。かかる翻訳モデル（原言語木構造モデル、目的言語木構造モデル、木構造マッピングモデルの総称）に含まれる規則は、階層的句アライメント方法（非特許文献３参照）等を用いると、コーパスから自動的に抽出することができる。また、これらモデルの確率は、コーパス中に規則が使われた頻度をカウントし、その相対頻度を計算するなどの処理により、算出することができる。かかる第三の技術により、原言語木構造モデル、目的言語木構造モデル、木構造マッピングモデルが自動的に取得でき得る。

【非特許文献1】Philipp Koehn, Franz J. Och, and Daniel Marcu:Statistical Phrase-Based Translation,HLT-NAACL 2003: Main Proceedings, 2003,pp.127-133

【非特許文献2】古瀬 蔵他２名，構成素境界解析を用いた多言語話し言葉翻訳，自然言語処理，Vol. 6, No. 5, 1999,pp.63-91

【非特許文献3】今村 賢治，構文解析と融合した階層的句アライメント，自然言語処理，Vol. 9, No. 5, 2002, pp. 23-42

【発明の開示】

【発明が解決しようとする課題】

【0005】

しかしながら、従来の第一の機械翻訳装置においては、句の順序を調整しなければ正しい翻訳文とはならない。本機械翻訳装置の翻訳方法では、句の順序調整を平坦な構造上で行い、言語モデルで検証していた。そのため、構文的に誤った翻訳文を出力することがある、という課題があった。

また、従来の第二の機械翻訳装置においては、入力の解析結果として、複数の構造が得られた時、単語の意味曖昧等を用いて曖昧性解消を行っていた。そのため、シソーラスを必要とする、という課題があった。

【課題を解決するための手段】

【0006】

本第一の発明の機械翻訳装置は、翻訳される元の文章の言語である第一言語の木構造に関する情報であり、複数の単語からなり非終端記号を含まない長単位句を一のノードとする木構造の情報を含む原言語木構造情報と、当該原言語木構造情報に対応する木構造の確率を示す情報である原言語木構造確率を有する原言語木構造レコードを１以上有する原言語木構造モデルを格納している原言語木構造モデル格納部と、翻訳結果の文章の言語である第二言語の木構造に関する情報であり、複数の単語からなり非終端記号を含まない長単位句を一のノードとする木構造の情報を含む目的言語木構造情報と、当該目的言語木構造情報に対応する木構造の確率を示す情報である目的言語木構造確率を有する目的言語木構造レコードを１以上有する目的言語木構造モデルを格納している目的言語木構造モデル格

納部と、原言語木構造情報と目的言語木構造情報との対応を示す情報であるマッピング情報と、当該マッピング情報が示す対応の確率を示す情報であるマッピング確率を有する木構造マッピングレコードを１以上有する木構造マッピングモデルを格納している木構造マッピングモデル格納部と、第二言語における単語の出現に関する確率の情報である単語出現確率を１以上有する言語モデルを格納している言語モデル格納部と、第一言語の文章を受け付ける受付部と、前記受付部が受け付けた文章を構文解析し、当該文章の一部または全部の木構造に関する情報である木構造情報を、順次得る入力文構文解析部と、前記入力文構文解析部が得た木構造情報に対応する１以上の原言語木構造確率を前記原言語木構造モデル格納部から取得し、前記入力文構文解析部が得た木構造情報に対応する１以上の目的言語木構造確率を前記目的言語木構造モデル格納部から取得し、前記入力文構文解析部が得た木構造情報に対応する１以上のマッピング確率を前記木構造マッピングモデル格納部から取得し、前記１以上の目的言語木構造情報のそれぞれに対応する１以上の目的言語木構造確率を前記目的言語木構造モデル格納部から取得し、前記取得した目的言語木構造情報に基づいて、当該目的言語木構造情報を構成する２以上の単語の、１以上の単語出現確率を前記言語モデル格納部から取得し、前記取得した原言語木構造確率、前記取得したマッピング確率、前記取得した目的言語木構造確率、および前記取得した単語出現確率に基づいて、出力の構文木の評価値を算出する木構造マッピング確率計算部と、前記木構造マッピング確率計算部が算出した評価値に基づいて、出力する第二言語の一部または全部の構文木を決定し、当該決定した構文木に基づいて第二言語の一部または全部の文章である出力情報を取得する最尤列探索部と、前記最尤列探索部が取得した１以上の出力情報を有する第二言語の文章を出力する出力部を具備する機械翻訳装置である。

かかる構成により、高品質かつ高速な翻訳が可能となる。

【0007】

また、本第二の発明の機械翻訳装置は、第一の発明に対して、前記木構造マッピング確率計算部は、前記入力文構文解析部が得た木構造情報に対応する１以上の原言語木構造確率を前記原言語木構造モデル格納部から取得する原言語木構造確率取得手段と、前記入力文構文解析部が得た木構造情報に対応する１以上の目的言語木構造情報と１以上のマッピング確率を前記木構造マッピングモデル格納部から取得するマッピング確率取得手段と、前記１以上の目的言語木構造情報のそれぞれに対応する１以上の目的言語木構造確率を前記目的言語木構造モデル格納部から取得する目的言語木構造確率取得手段と、前記取得した目的言語木構造情報に基づいて、当該目的言語木構造情報を構成する２以上の単語の、１以上の単語出現確率を前記言語モデル格納部から取得する単語出現確率取得手段と、前記原言語木構造確率、前記マッピング確率、前記目的言語木構造確率、および前記単語出現確率に基づいて、出力の構文木の評価値を算出する評価値算出手段を具備する機械翻訳装置である。

かかる構成により、高品質かつ高速な翻訳が可能となる。

【0008】

また、本第三の発明の機械翻訳装置は、第一、第二の発明に対して、前記木構造マッピング確率計算部が算出した２以上の評価値と当該評価値に対応する出力の構文木において、同一の構文木に対応する評価値を合成するマージ部をさらに具備し、前記最尤列探索部は、前記木構造マッピング確率計算部が算出した評価値および前記マージ部が合成して得た評価値に基づいて、出力する第二言語の一部または全部の構文木を決定し、当該決定した構文木に基づいて第二言語の一部または全部の文章である出力情報を取得する機械翻訳装置である。

かかる構成により、高品質かつ高速な翻訳が可能となる。

また、本第四の発明の機械翻訳装置は、第二、第三の発明に対して、前記評価値算出手段は、前記原言語木構造確率、前記マッピング確率、前記目的言語木構造確率、前記単語出現確率の積を算出し、当該積を出力の構文木の評価値とする機械翻訳装置である。

かかる構成により、高品質かつ高速な翻訳が可能となる。

【0009】

また、本第五の発明の機械翻訳装置は、第一から第四いずれかの発明に対して、前記最

尤列探索部は、前記木構造マッピング確率計算部が算出した評価値が最大の構文木を、出力する構文木として決定し、当該決定した構文木に基づいて第二言語の一部または全部の文章である出力情報を取得する機械翻訳装置である。

かかる構成により、高品質かつ高速な翻訳が可能となる。

【発明の効果】

【0010】

本発明による機械翻訳装置によれば、高品質かつ高速な翻訳が可能となる。

【発明を実施するための最良の形態】

【0011】

以下、機械翻訳装置等の実施形態について図面を参照して説明する。なお、実施の形態において同じ符号を付した構成要素は同様の動作を行うので、再度の説明を省略する場合がある。

（実施の形態１）

図１は、本実施の形態における機械翻訳装置のブロック図である。

【0012】

機械翻訳装置は、原言語木構造モデル格納部１１、目的言語木構造モデル格納部１２、木構造マッピングモデル格納部１３、言語モデル格納部１４、受付部１５、入力文構文解析部１６、木構造マッピング確率計算部１７、マージ部１８、最尤列探索部１９、出力部２０を具備する。また、機械翻訳装置は、入力手段として、例えば、キーボード３０２、マウス３０３を具備する。さらに、機械翻訳装置は、出力手段として、例えば、ディスプレイ３０４を具備する。

木構造マッピング確率計算部１７は、原言語木構造確率取得手段１７１、マッピング確率取得手段１７２、目的言語木構造確率取得手段１７３、単語出現確率取得手段１７４、評価値算出手段１７５を具備する。

【0013】

原言語木構造モデル格納部１１は、翻訳される元の文章の言語である第一言語の木構造に関する情報であり、１つの非終端信号に対し、１個以上の終端信号または非終端信号を対応付けた木構造の情報を含む原言語木構造情報と、当該原言語木構造情報に対応する木構造の確率を示す情報である原言語木構造確率を有する原言語木構造レコードを１以上有する原言語木構造モデルを格納している。なお、非終端記号とは、構文ラベル付き変数である。終端記号とは、単語そのものである。また、原言語木構造情報は、以下の複数の種類がある。例えば、原言語木構造情報は、２個以上の終端記号を子ノードとして持ち、親ノードが非終端記号である規則を含む情報である原言語長単位句木構造情報と、１つ以上の非終端記号と、０個以上の終端記号を子ノードとして持つ、親ノードが非終端記号である規則を示す情報である原言語構文木構造情報と、１個の終端記号を子ノードとして持ち、親ノードが非終端記号である規則を示す情報である原言語単語単位木構造情報などがある。また、原言語木構造確率とは、原言語において、親ノードの非終端記号が、子ノードの非終端記号列または終端記号列を生成する確率である。また、１つの非終端信号に対し、１個以上の終端信号または非終端信号を対応付けた木構造の情報は、複数の単語からなり非終端記号を含まない長単位句を一のノードとする木構造の情報を含む。原言語木構造モデルのデータ構造は問わない。原言語木構造モデルの例は、後述する。原言語木構造モデル格納部１１は、不揮発性の記録媒体が好適であるが、揮発性の記録媒体でも実現可能である。

【0014】

目的言語木構造モデル格納部１２は、翻訳結果の文章の言語である第二言語の木構造に関する情報であり、１つの非終端信号に対し、１個以上の終端信号または非終端信号を対応付けた木構造の情報を含む目的言語木構造情報と、当該目的言語木構造情報に対応する木構造の確率を示す情報である目的言語木構造確率を有する目的言語木構造レコードを１以上有する目的言語木構造モデルを格納している。目的言語木構造情報は、以下の複数の種類がある。例えば、目的言語木構造情報は、２個以上の終端記号を子ノードとして持ち

、親ノードが非終端記号である規則を含む情報である目的言語長単位句木構造情報と、１つ以上の非終端記号と、０個以上の終端記号を子ノードとして持つ、親ノードが非終端記号である規則を示す情報である目的言語構文木構造情報と、１個の終端記号を子ノードとして持ち、親ノードが非終端記号である規則を示す情報である目的言語単語単位木構造情報などがある。また、目的言語木構造確率とは、目的言語において、親ノードの非終端記号が、子ノードの非終端記号列または終端記号列を生成する確率である。また、１つの非終端信号に対し、１個以上の終端信号または非終端信号を対応付けた木構造の情報は、複数の単語からなり非終端記号を含まない長単位句を一のノードとする木構造の情報を含む。目的言語木構造モデルのデータ構造は問わない。目的言語木構造モデルの例は、後述する。目的言語木構造モデル格納部１２は、不揮発性の記録媒体が好適であるが、揮発性の記録媒体でも実現可能である。

【0015】

木構造マッピングモデル格納部１３は、原言語木構造情報と目的言語木構造情報との対応を示す情報であるマッピング情報と、当該マッピング情報が示す対応の確率を示す情報であるマッピング確率を有する木構造マッピングレコードを１以上有する木構造マッピングモデルを格納している。ここでは、マッピング確率とは、原言語木構造情報と目的言語木構造情報が対応する確率を示す情報である。木構造マッピングモデルのデータ構造は問わない。木構造マッピングモデルの例は、後述する。木構造マッピングモデル格納部１３は、不揮発性の記録媒体が好適であるが、揮発性の記録媒体でも実現可能である。

【0016】

言語モデル格納部１４は、第二言語における単語の出現に関する確率の情報である単語出現確率を１以上有する言語モデルを格納している。単語出現確率は、例えば、第二言語の、２以上の単語の連続した出現に関する確率の情報である。単語出現確率は、例えば、第二言語の第一の単語と、第二の単語と、第一の単語の次に第二の単語が出現する確率を示す情報である。言語モデルは、例えば、第一の単語と、第二の単語と、単語出現確率を有するレコードを1以上有する。言語モデル格納部１４は、不揮発性の記録媒体が好適であるが、揮発性の記録媒体でも実現可能である。

【0017】

受付部１５は、第一言語の文章を受け付ける。文章の入力手段は、テンキーやキーボードやマウスやメニュー画面によるもの等、何でも良い。受付部１５は、テンキーやキーボード等の入力手段のデバイスドライバーや、メニュー画面の制御ソフトウェア等で実現され得る。

【0018】

入力文構文解析部１６は、受付部１５が受け付けた文章を構文解析し、当該文章の一部または全部の木構造に関する情報である木構造情報を、順次得る。入力文構文解析部１６は、通常、原言語木構造モデル格納部１１の原言語木構造モデルを用いて、文章を構文解析する。ただし、入力文構文解析部１６は、原言語木構造モデルを用いることが好適であるが、他の木構造モデルを用いて、文章を構文解析しても良いし、他の手段により文章を構文解析しても良い。なお、文章を構文解析し、木構造情報を、順次得る処理は公知技術における処理であるので、詳細な説明は省略する。入力文構文解析部１６は、通常、ＭＰＵやメモリ等から実現され得る。入力文構文解析部１６の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはＲＯＭ等の記録媒体に記録されている。但し、ハードウェア（専用回路）で実現しても良い。

【0019】

木構造マッピング確率計算部１７は、入力文構文解析部１６が得た木構造情報に対応する１以上の原言語木構造確率を原言語木構造モデル格納部１１から取得し、入力文構文解析部１６が得た木構造情報に対応する１以上の目的言語木構造情報と１以上のマッピング確率を木構造マッピングモデル格納部１３から取得し、１以上の目的言語木構造情報のそれぞれに対応する１以上の目的言語木構造確率を目的言語木構造モデル格納部１２から取得し、取得した目的言語木構造情報に基づいて、当該目的言語木構造情報を構成する２以

上の単語の、１以上の単語出現確率を言語モデル格納部１４から取得し、取得した原言語木構造確率、取得したマッピング確率、取得した目的言語木構造確率、および取得した単語出現確率に基づいて、出力の構文木の評価値を算出する。木構造マッピング確率計算部１７は、通常、ＭＰＵやメモリ等から実現され得る。木構造マッピング確率計算部１７の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはＲＯＭ等の記録媒体に記録されている。但し、ハードウェア（専用回路）で実現しても良い。

原言語木構造確率取得手段１７１は、入力文構文解析部１６が得た木構造情報に対応する１以上の原言語木構造確率を原言語木構造モデル格納部１１から取得する。

マッピング確率取得手段１７２は、入力文構文解析部１６が得た木構造情報に対応する１以上の目的言語木構造情報と１以上のマッピング確率を木構造マッピングモデル格納部１３から取得する。

【0020】

目的言語木構造確率取得手段１７３は、１以上の目的言語木構造情報のそれぞれに対応する１以上の目的言語木構造確率を目的言語木構造モデル格納部１２から取得する。ここでの１以上の目的言語木構造情報は、マッピング確率取得手段１７２が取得した目的言語木構造情報である。

【0021】

単語出現確率取得手段１７４は、取得した目的言語木構造情報に基づいて、当該目的言語木構造情報を構成する２以上の単語の、１以上の単語出現確率を言語モデル格納部１４から取得する。単語出現確率取得手段１７４は、例えば、後述する単語ｂｉｇｒａｍモデルの確率を取得しても良い。

【0022】

評価値算出手段１７５は、原言語木構造確率、マッピング確率、目的言語木構造確率、および単語出現確率に基づいて、出力の構文木の評価値を算出する。評価値算出手段１７５は、原言語木構造確率、マッピング確率、目的言語木構造確率の積で翻訳モデル確率を算出し、かつ、１以上の単語出現確率の積で言語モデル確率を算出し、かつ当該翻訳モデル確率と言語モデル確率に基づいて評価値を算出することは好適である。さらに、評価値算出手段１７５は、翻訳モデル確率と言語モデル確率の積により評価値を算出することは好適である。

【0023】

原言語木構造確率取得手段１７１、マッピング確率取得手段１７２、目的言語木構造確率取得手段１７３、単語出現確率取得手段１７４、評価値算出手段１７５は、通常、ＭＰＵやメモリ等から実現され得る。原言語木構造確率取得手段１７１等の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはＲＯＭ等の記録媒体に記録されている。但し、ハードウェア（専用回路）で実現しても良い。

マージ部１８は、木構造マッピング確率計算部１７が算出した２以上の評価値と当該評価値に対応する出力の構文木において、同一の構文木に対応する評価値を合成する。合成とは、例えば、２以上の評価値の和を算出することである。また、合成とは、例えば、一の出力の構文木に対応する２以上の評価値を、一の評価値としてグループ化（リンク付けなど）することである。具体的には、マージ部１８は、例えば、得られた出力の構文木の木構造情報（例えば、単語列リスト）をキーとして、バッファ（例えば、出力の単語列リスト、および確率が一時格納されたバッファ）を検索する。なお、かかるバッファには、木構造マッピング確率計算部１７が得た、出力の木構造情報（単語列リスト）と評価値（例えば、確率）が格納されている。そして、マージ部１８は、バッファ中に、得られた木構造情報（例えば、単語列リスト）が存在するか否かを判断する。そして、マージ部１８は、得られた木構造情報（例えば、単語列リスト）が存在すると判断した場合、得られた木構造情報（例えば、単語列リスト）に対応する評価値（例えば、確率）として、得られた評価値（例えば、確率）を追記する。なお、一の単語リストに対応する評価値（例えば、確率）が２以上存在する場合、翻訳モデル確率（評価値）の和が、当該単語リストの翻訳確率（評価値）となる。そして、例えば、「（２つ以上の翻訳モデル確率の和）×言語

モデル確率」が当該単語リストの確率（評価値）となる。そして、マージ部１８は、得られた木構造情報（例えば、単語列リスト）が存在しないと判断した場合、得られた木構造情報（例えば、単語列リスト）、および得られた評価値（例えば、確率）を対にして登録する。マージ部１８は、通常、ＭＰＵやメモリ等から実現され得る。マージ部１８の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはＲＯＭ等の記録媒体に記録されている。但し、ハードウェア（専用回路）で実現しても良い。

【0024】

最尤列探索部１９は、木構造マッピング確率計算部１７が算出した評価値に基づいて、出力する第二言語の一部または全部の構文木を決定し、当該決定した構文木に基づいて第二言語の一部または全部の文章である出力情報を取得する。最尤列探索部１９は、木構造マッピング確率計算部１７が算出した評価値が最大の構文木を出力する構文木として決定し、当該決定した構文木に基づいて第二言語の一部または全部の文章である出力情報を取得することは好適である。最尤列探索部１９は、通常、ＭＰＵやメモリ等から実現され得る。最尤列探索部１９の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはＲＯＭ等の記録媒体に記録されている。但し、ハードウェア（専用回路）で実現しても良い。

【0025】

出力部２０は、最尤列探索部１９が取得した１以上の出力情報を有する第二言語の文章を出力する。ここで、出力とは、ディスプレイへの表示、プリンタへの印字、音声合成による出力、外部の装置への送信等を含む概念である。出力部２０は、ディスプレイやスピーカー等の出力デバイスを含むと考えても含まないと考えても良い。出力部２０は、出力デバイスのドライバーソフトまたは、出力デバイスのドライバーソフトと出力デバイス等で実現され得る。

なお、第一言語、第二言語は、日本語、英語、中国語、韓国語等何でも良い。ただし、第一言語と第二言語は異なる言語である。

次に、機械翻訳装置の動作について図２から図５のフローチャートを用いて説明する。

（ステップＳ２０１）受付部１５は、第一言語の文章を受け付けたか否かを判断する。文章を受け付ければステップＳ２０２に行き、文章を受け付けなければステップＳ２０１に戻る。

【0026】

（ステップＳ２０２）入力文構文解析部１６は、原言語木構造モデル格納部１１の原言語木構造モデル（詳細には、原言語木構造情報）を用いて、ステップＳ２０１で受け付けた文章を構文解析し、当該文章の一部または全部の木構造に関する情報である木構造情報を得る。なお、本ステップにおける文章の構文解析処理は、先に読み込んだ単語を次々に用いて、ボトムアップに大きな構文木を構成するような処理である。

【0027】

（ステップＳ２０３）木構造マッピング確率計算部１７は、ステップＳ２０２において新たな部分木（新たな木構造情報）が取得できたか否かを判断する。新たな部分木が取得できればステップＳ２０４に行き、新たな部分木が取得できなければステップＳ２０５に行く。

【0028】

（ステップＳ２０４）木構造マッピング確率計算部１７は、評価処理を行う。評価処理とは、出力の構文木の評価値を算出する処理である。出力の構文木の評価値は、出力情報（翻訳された文章）が出力される確率を示す情報である。ステップＳ２０２に戻る。なお、ステップＳ２０２に戻った際には、ステップＳ２０２において、直前に作成した部分木を構成する前の単語をも取得して、ボトムアップにより大きな構文木を構成するように処理される。評価処理の詳細について、図３のフローチャートを用いて説明する。

【0029】

（ステップＳ２０５）最尤列探索部１９は、文章の最後尾か否かを判断する。最後尾であればステップＳ２０６に行き、最後尾でなければステップＳ２１０に行く。なお、ステ

ップＳ２０５における判断が最後尾でないとの判断の場合、本文章は、２以上の部分木を有することとなる。

【0030】

（ステップＳ２０６）最尤列探索部１９は、上述の構文解析の処理、および評価処理において、バッファ上に登録された複数の翻訳候補（第二言語の１以上の構文木の集合）から、部分木（構文木）の数が最少となる部分木列を取得する。

【0031】

（ステップＳ２０７）最尤列探索部１９は、各部分木の出力単語リストから、確率の和が最大となる単語列を決定する。その際、同一の出力単語リストが２以上存在する場合には、当該出力単語リストに対応する確率の和を算出し、確率の和が最大となる単語列を決定する際に、当該確率の和を比較対象とする。

（ステップＳ２０８）最尤列探索部１９は、ステップＳ２０６で決定した１以上の部分木の単語列を連結する。

【0032】

（ステップＳ２０９）出力部２０は、ステップＳ２０８で連結された単語列を出力する。なお、かかる単語列は、最尤列探索部１９が取得した１以上の出力情報を有する第二言語の文章（翻訳結果）である。処理を終了する。

（ステップＳ２１０）最尤列探索部１９は、次の単語にスキップする。次の単語へのスキップとは、構文解析を行う単語のポインタをずらす処理である。ステップＳ２０２に戻る。

次に、ステップＳ２０４の評価処理について、図３のフローチャートを用いて詳細に説明する。

【0033】

（ステップＳ３０１）木構造マッピング確率計算部１７は、構築した部分木の最上位ノードの規則θ_fに対応するすべての目的言語木構造モデルの規則θ_eを、目的言語木構造モデル格納部１２から取得する。つまり、ここでは、１以上の（θ_f，θ_e）の組が取得される。なお、規則とは、原言語木構造情報、目的言語木構造情報等である。

（ステップＳ３０２）木構造マッピング確率計算部１７は、カウンタｉに１を代入する。

【0034】

（ステップＳ３０３）木構造マッピング確率計算部１７は、ｉ番目の（θ_f，θ_e）の組が存在するか否かを判断する。ｉ番目の（θ_f，θ_e）の組が存在すればステップＳ３０４に行き、存在しなければ上位関数にリターンする。

【0035】

（ステップＳ３０４）木構造マッピング確率計算部１７は、ｉ番目の（θ_f，θ_e）の組のθ_eを用いて、出力の部分木（目的言語木構造情報）を構築する。木構造マッピング確率計算部１７は、以下のように出力の部分木を構築する。つまり、入力文構文解析部１６により入力文をボトムアップに構文解析しているので、入力構文木におけるθ_fの子ノードの非終端記号は既にわかっている。すると、θ_eの子ノードの非終端記号についても、出力構文木（と出力単語列リスト）は既にわかっていることになる。θ_eの子ノードの出力構文木を、θ_eの非終端記号に埋め込む際、子ノードの出力構文木の最上位の構文ラベルと、非終端記号の構文ラベルを比較し、θ_eのすべての子ノード非終端記号について一致している場合だけ埋め込んで、θ_eの親ノードをトップとする出力構文木を、木構造マッピング確率計算部１７は生成する。

（ステップＳ３０５）木構造マッピング確率計算部１７は、ステップＳ３０４において、出力の部分木が構築できたか否かを判断する。

（ステップＳ３０６）木構造マッピング確率計算部１７は、構築できた部分木を単語列リストに展開する。

（ステップＳ３０７）木構造マッピング確率計算部１７は、当該部分木の翻訳モデル確率を算出する。翻訳モデル確率を算出する処理については、図４のフローチャートを用い

て、詳細に説明する。

（ステップＳ３０８）木構造マッピング確率計算部１７は、当該部分木の言語モデル確率を算出する。言語モデル確率を算出する処理については、図５のフローチャートを用いて、詳細に説明する。

【0036】

（ステップＳ３０９）木構造マッピング確率計算部１７は、ステップＳ３０７で算出した翻訳モデル確率と、ステップＳ３０８で算出した言語モデル確率を用いて、出力単語列の確率を算出する。通常、木構造マッピング確率計算部１７は、「翻訳モデル確率×言語モデル確率」により、出力単語列の確率を算出する。出力単語列の確率とは、入力単語列が出力単語列に翻訳される確率である。

（ステップＳ３１０）マージ部１８は、ステップＳ３０６で得られた単語列リストをキーとして、バッファ（出力の単語列リスト、および確率が一時格納されたバッファ）を検索する。

【0037】

（ステップＳ３１１）マージ部１８は、バッファ中に、ステップＳ３０６で得られた単語列リストが存在するか否かを判断する。単語列リストが存在すればステップＳ３１２に行き、単語列リストが存在しなければステップＳ３１３に行く。

【0038】

（ステップＳ３１２）マージ部１８は、ステップＳ３０６で得られた単語列リストに対応する確率として、ステップＳ３０９で得られた確率を追記する。ステップＳ３１４に行く。なお、一の単語リストに対応する確率が２以上存在する場合、翻訳モデル確率の和が、該当単語リストの翻訳確率となる。そして、「（２つ以上の翻訳モデル確率の和）×言語モデル確率」が当該単語リストの確率となる。

（ステップＳ３１３）マージ部１８は、ステップＳ３０６で得られた単語列リスト、およびステップＳ３０９で得られた確率を対にして登録する。

（ステップＳ３１４）木構造マッピング確率計算部１７は、カウンタｉを１、インクリメントする。ステップＳ３０３に戻る。

次に、ステップＳ３０７の翻訳モデル確率を算出する処理について、図４のフローチャートを用いて、詳細に説明する。

（ステップＳ４０１）原言語木構造確率取得手段１７１は、原言語の木構造を用いて、原言語木構造確率を、原言語木構造モデル格納部１１から取得する。

【0039】

（ステップＳ４０２）マッピング確率取得手段１７２は、原言語の木構造（原言語木構造情報）、および出力の部分木（目的言語木構造情報）を用いて、マッピング確率を木構造マッピングモデル格納部１３から取得する。

（ステップＳ４０３）目的言語木構造確率取得手段１７３は、出力の部分木を用いて、目的言語木構造確率を目的言語木構造モデル格納部１２から取得する。

【0040】

（ステップＳ４０４）評価値算出手段１７５は、原言語木構造確率、マッピング確率、および目的言語木構造確率に基づいて、出力の構文木の翻訳モデル確率を算出する。評価値算出手段１７５は、原言語木構造確率、マッピング確率、目的言語木構造確率の積で翻訳モデル確率を算出することは好適である。上位関数にリターンする。

なお、図４のフローチャートにおいて、原言語木構造確率、マッピング確率、目的言語木構造確率を取得する順序は問わないことは言うまでもない。

【0041】

次に、ステップＳ３０８の言語モデル確率を算出する処理について、図５のフローチャートを用いて、詳細に説明する。本フローチャートにおいて算出する言語モデル確率の言語モデルは、単語ｂｉｇｒａｍモデルである。

（ステップＳ５０１）木構造マッピング確率計算部１７は、カウンタｉに１を代入する。

（ステップＳ５０２）木構造マッピング確率計算部１７は、ｉ番目の単語が存在するか否かを判断する。ｉ番目の単語が存在すればステップＳ５０３に行き、ｉ番目の単語が存在しなければステップＳ５０４に行く。

【0042】

（ステップＳ５０３）木構造マッピング確率計算部１７は、（ｉ－１）番目の単語、ｉ番目の単語を取得する。なお、ｉが「１」の時は、木構造マッピング確率計算部１７は、「＜Ｓ＞」および１番目の単語を取得する。「＜Ｓ＞」は、文の開始を示す記号の文開始記号である。ステップＳ５０５に行く。

（ステップＳ５０４）木構造マッピング確率計算部１７は、（ｉ－１）番目の単語、「＜／Ｓ＞」を取得する。なお、「＜／Ｓ＞」は、文の終了を示す記号の文終了記号である。

【0043】

（ステップＳ５０５）単語出現確率取得手段１７４は、「＜Ｓ＞、ｉ番目の単語」、「（ｉ－１）番目の単語、ｉ番目の単語」または「（ｉ－１）番目の単語、＜／Ｓ＞」に対応する情報を、言語モデル格納部１４から検索する。

【0044】

（ステップＳ５０６）単語出現確率取得手段１７４は、「＜Ｓ＞、ｉ番目の単語」、「（ｉ－１）番目の単語、ｉ番目の単語」または「（ｉ－１）番目の単語、＜／Ｓ＞」が言語モデル格納部１４に存在するか否かを判断する。言語モデル格納部１４に存在すればステップＳ５０７に行き、存在しなければステップＳ５０８に行く。

（ステップＳ５０７）単語出現確率取得手段１７４は、対応する単語出現確率を、言語モデル格納部１４から取得し、一時蓄積する。ステップＳ５０９に行く。

（ステップＳ５０８）単語出現確率取得手段１７４は、単語出現確率を予め決められた値とし、一時蓄積する。なお、予め決められた値は、単語出現確率取得手段１７４が保持している、とする。

（ステップＳ５０９）木構造マッピング確率計算部１７は、カウンタｉを１、インクリメントする。

【0045】

（ステップＳ５１０）木構造マッピング確率計算部１７は、ラストか否かを判断する。ラストであればステップＳ５１１に行き、ラストでなければステップＳ５０２に戻る。なお、ラストか否かは、「＜／Ｓ＞」が出現したか否かにより判断され得る。

（ステップＳ５１１）木構造マッピング確率計算部１７は、一時蓄積した１以上の単語出現確率の積を算出する。かかる１以上の単語出現確率の積が、言語モデル確率である。上位関数にリターンする。

【0046】

なお、図５のフローチャートにおいて、言語モデルは、単語ｂｉｇｒａｍモデルを用いたが、単語ｔｒｉｇｒａｍモデル、品詞ｔｒｉｇｒａｍモデルなどの、他の言語モデルを用いても良い。他の言語モデルについては、公知技術であるので、詳細な説明は省略する。

上記のフローチャートで説明した機械翻訳装置の翻訳方法は、以下の翻訳方法である。

【0047】

つまり、本翻訳方法は、統計翻訳の一種である。統計翻訳は、入力の単語列ｆが与えられたとき、確率を最大化する出力の単語列ｅを、すべての可能な組み合わせ中から探索することにより、翻訳を行う方法である。探索結果は、以下の数式１により表わされる。数式１において「ａｒｇｍａｘ」は、確率を最大化する出力の単語列を取得することを示す。確率を最大化する出力の単語列とは、翻訳結果の文章（第二言語の文章）である。

【数1】

$$\hat{e} = \underset{e}{\operatorname{argmax}} P(e|f)$$

$$= \underset{e}{\operatorname{argmax}} P(e|f)^2$$

$$= \underset{e}{\operatorname{argmax}} P(e)P(f|e)P(e|f). \quad (\hat{e}は探索結果)$$

なお、数式１において、Ｐ（ｅ）を言語モデル確率、Ｐ（ｆ｜ｅ）を逆方向翻訳モデル確率、Ｐ（ｅ｜ｆ）を順方向翻訳モデル確率、Ｐ（ｆ｜ｅ）Ｐ（ｅ｜ｆ）を単に翻訳モデル確率と言う。

つまり、本具体例において、翻訳結果の文章を取得することは、言語モデル確率と翻訳モデル確率の積が最大の単語列を取得することである。

以下、翻訳モデル確率の算出方法について、数式を用いて説明する。

【0048】

構文トランスファ方式の統計翻訳は、翻訳モデル中に隠れ変数として原言語・目的言語の構文木（それぞれF、Ｅと示し、単語列ｆ，ｅを生成する）を仮定し、木構造同士のマッピングを行うことにより、翻訳文を生成する。

木構造翻訳描置において、翻訳モデルを原言語(入力)の木構造モデル、目的言語(出力)の木構造モデル、および順方向・逆方向木構造マッピングモデルに分解する。

具体的には、数式２で表現される。

【数2】

$$P(e|f)P(f|e) = \sum_{E,F} P(E|F)P(F|f) \sum_{F,E} P(F|E)P(E|e)$$

$$\simeq \sum_{F,E} P(F|f)P(E|F)P(F|E)P(E|e).$$

【0049】

ここで、Ｐ（Ｆ｜ｆ）は、原言語の木構造モデル確率である。また、Ｐ（Ｅ｜ｅ）は、目的言語の木構造モデル確率である。また、Ｐ（Ｅ｜Ｆ）は、順方向木構造マッピングモデル確率である。また、Ｐ（Ｆ｜Ｅ）は、逆方向木構造マッピングモデル確率である。Ｐ（Ｅ｜Ｆ）Ｐ（Ｆ｜Ｅ）を、単に木構造マッピングモデル確率と言う。

【0050】

しかし、構文木全体を変換することはできないため、構文木を構成する文脈自由文法規則単位に変換を行う。文脈自由文法規則単位とは、親ノードの非終端記号に対して、子ノードの非終端記号または終端記号列を生成する規則である。たとえば、原言語の構文木Ｅを構成する文脈自由文法規則をθ_e、目的言語の構文木Ｆ構成する文脈自由文法規則をθ_fとしたとき、各モデルの確率は、以下の数式３から数式５で算出する。数式３は、原言語の木構造モデル確率を算出する式である。数式４は、目的言語の木構造モデル確率を算出する式である。数式５は、木構造マッピングモデル確率を算出する式である。

【数3】

$$P(F|f) = \prod_{\theta_f;\ \theta_f \in F} P(\theta_f),$$

【数4】

$$P(E|e) = \prod_{\theta_e;\ \theta_e \in E} P(\theta_e),$$

【数5】

$$P(E|F)\,P(F|E) \;=\; \prod_{\substack{\theta_e,\,\theta_f\in F\\ \theta_e:\,\theta_e\in E}} P(\theta_e|\,\theta_f)\,P(\theta_f|\,\theta_e).$$

【0051】

なお、本機械翻訳装置において、翻訳を行う際には、翻訳モデル確率は再帰的に計算する。たとえば、ある部分木の最上位ノードＮｉが、その直下にｊ個の部分木を含んでいる場合、以下の数式６で算出する。

【数6】

$$P(F^i|f^i)P(E^i|F^i)P(F^i|E^i)P(E^i|e^i)$$

$$= \prod_{\substack{\theta^k_f:\,\theta^k_f\in F^i\\ \theta^k_e:\,\theta^k_e\in E^i}} P(\theta^k_e)P(\theta^k_f)P(\theta^k_f|\theta^k_e)P(\theta^k_e|\theta^k_f)$$

$$= P(\theta^i_e)P(\theta^i_f|\theta^i_e)P(\theta^i_e|\theta^i_f) \cdot \prod_{k=1}^{j} P(F^{i+k}|f^{i+k})P(E^{i+k}|F^{i+k})P(F^{i+k}|E^{i+k})P(E^{i+k}|e^{i+k})$$

【数7】

$$P(e) \;=\; \prod_{i=1}^{m+1} P(w^i_e|\,w^{i-1}_e),$$

（ただし、w^i_eはeの単語で、$e = w^1_e \cdots w^m_e$である。また、w^0_eは文開始記号（＜s＞と表記）、w^{m+1}_eは、文終了記号（＜/s＞と表記）を意味する。）

【0052】

本機械翻訳装置において、上述した数式により、第二言語の文章（出力単語列リスト）を評価し、例えば、最も評価値が大きい出力単語列リストを出力する。かかる出力単語列リストが、翻訳結果の文章（第二言語の文章）である。

以下、本実施の形態における機械翻訳装置の具体的な動作について説明する。

【0053】

図６は、本機械翻訳装置の原言語木構造モデル格納部１１に格納されている原言語木構造モデルの例である。本原言語木構造モデルは、名前、原言語木構造情報、原言語木構造確率「Ｐ（θ_f）」の属性値を有する１以上の原言語木構造レコードを有する。名前は、原言語木構造情報、原言語木構造確率の組を識別する情報である。原言語木構造情報は、親ノードと子ノード列を有する。つまり、原言語木構造情報は、親ノードと子ノード列により、木構造を構成している。また、原言語木構造情報の中には、２個以上の終端記号を子ノードとして持ち、親ノードが非終端記号である規則を含む情報である原言語長単位句木構造情報が含まれる。原言語長単位句木構造情報は、例えば、名前「ＳＲＣ－１０２」、「ＳＲＣ－１０３」、「ＳＲＣ－１０５」に対応する原言語木構造情報である。また、図６の原言語木構造確率「Ｐ（θ_f）」等における、例えば、「３．６７ｅ－３」は、「$３．６７\times１０^{-３}$」のことである。また、原言語木構造確率「Ｐ（θ_f）」は、原言語において、親ノードが子ノード列を生成する確率である。

【0054】

図７は、本機械翻訳装置の目的言語木構造モデル格納部１２に格納されている目的言語木構造モデルの例である。本目的言語木構造モデルは、名前、目的言語木構造情報、目的

言語木構造確率「Ｐ（θ_e）」の属性値を有する１以上の目的言語木構造レコードを有する。名前は、目的言語木構造情報、目的言語木構造確率の組を識別する情報である。目的言語木構造情報は、原言語木構造情報と同様に、親ノードと子ノード列を有する。また、目的言語木構造確率「Ｐ（θ_e）」は、目的言語において、対応する目的言語木構造情報が現れる確率である。

【００５５】

図８は、木構造マッピングモデル格納部１３に格納されている木構造マッピングモデルの例である。本木構造マッピングモデルは、マッピング情報とマッピング確率「Ｐ（θ_e｜θ_f）Ｐ（θ_f｜θ_e）」を有する木構造マッピングレコードを１以上有する。マッピング情報は、目的言語の名前と原言語の名前の情報を有する。つまり、マッピング情報は、２つの名前で特定される目的言語の木構造と、原言語の木構造との対応を示す情報である。また、マッピング確率「Ｐ（θ_e｜θ_f）Ｐ（θ_f｜θ_e）」は、順方向の木構造マッピングモデル確率と逆方向の木構造マッピングモデル確率の積である。

【００５６】

図９は、言語モデル格納部１４に格納されている言語モデルの例である。本言語モデルは、第一の単語（w_e^{i-1}）と、第二の単語（w_e^i）と、第一の単語の次に第二の単語が出現する確率を示す情報である単語出現確率（Ｐ（w_e^i｜w_e^{i-1}））を有する単語出現確率を１以上保持している。なお、図９において「＜Ｓ＞」は文開始記号、「＜／Ｓ＞」は文終了記号である。

【００５７】

かかる状況において、本機械翻訳装置は、例えば、図１０に示すような構文トランスファ方式の機械翻訳を行う。また、本機械翻訳装置は、ここでは、日英翻訳を行う、とする。図１０は、「バスは１２時に出ますか」という日本語の文章を構文解析し、日本語構文木Ｆを取得し、次に、当該日本語構文木Ｆと英語構文木Ｅとのマッピングを行って英語文を出力することを示している。なお、図１０において、「Ｓ」は文章、「ＮＰ」は名詞句、「ＮＯＵＮ」は名詞、「ＶＰ」は動詞句、「ＰＰ」は副詞句、「Ｖ」は動詞、「Ｐ」は助詞、「ＮＵＭ」は数字を示す。また、図１０において、「ＳＱ」は疑問文、「ＮＮ」は名詞、「ＶＢ」は動詞、「ＩＮ」は前置詞、「ＣＤ」は数字を示す。（正しいでしょうか？）

本機械翻訳装置は、かかる構文トランスファ方式の機械翻訳において、翻訳モデル中に、複数の単語から成り、非終端記号を含まない規則（長単位句木構造情報）を含んでいる。

そして、ユーザは、例えば、キーボードから第一言語の文章「バスは１２時に出ますか」を入力した、とする。そして、受付部１５は、第一言語の文章「バスは１２時に出ますか」を受け付ける。

以下、かかる場合において、本機械翻訳装置が、「１２時に出」の部分翻訳を行う際の処理について説明する。

そして、本機械翻訳装置は、図１１に示す長単位句木構造情報、およびそれぞれに対応する翻訳モデル確率、言語モデル確率を保持している、とする。

【００５８】

次に、入力文構文解析部１６は、図６の原言語木構造モデル（詳細には、原言語木構造情報）を用いて、「１２時に出」を構文解析する。そして、入力文構文解析部１６は、ＳＲＣ－００２，ＳＲＣ－００３，ＳＲＣ－１０３が適用可能であり、どれも「１２時に出」についてＶＰを構成することを検出し、ＳＲＣ－００２，ＳＲＣ－００３，ＳＲＣ－１０３を取得する。かかる３つの部分木「ＳＲＣ－００２」，「ＳＲＣ－００３」，「ＳＲＣ－１０３」は、入力部分木を示す情報（その名前）である。

そして、木構造マッピング確率計算部１７は、３つの入力部分木（原言語木構造情報）に対して、以下の処理を行う。

【００５９】

まず、木構造マッピング確率計算部１７は、入力の部分木の最上位ノードに使われた規

則θ_f（ここでは、原言語木構造モデルの名前で特定する）に対応する目的言語木構造モデルの規則θ_e。（ここでは、目的言語木構造モデルの名前で特定する）を図８の木構造マッピングモデルから取得する。ＳＲＣ－００２に対応するθ_eは、ＴＲＧ－００２，ＴＲＧ－００３、ＳＲＣ－００３に対応するθ_eはＴＲＧ－００２，ＴＲＧ－００３、ＳＲＣ－１０３に対応するθ_eはＴＲＧ－１０４である。

【００６０】

ここで、木構造マッピング確率計算部１７は、（θ_f，θ_e）の組「（ＳＲＣ－００２，ＴＲＧ－００２）」「（ＳＲＣ－００２，ＴＲＧ－００３）」「（ＳＲＣ－００３，ＴＲＧ－００２）」「（ＳＲＣ－００３，ＴＲＧ－００３）」「（ＳＲＣ－１０３，ＴＲＧ－１０４）」を取得する。

次に、木構造マッピング確率計算部１７は、上記５組の（θ_f，θ_e）のそれぞれに対して、以下の処理を行う。

【００６１】

まず、木構造マッピング確率計算部１７は、θ_eを用いて、部分木を構築する（図１２のステップ１、ステップ２）。この例では、図１２の３つの部分木が構築される。（ＳＲＣ－００２，ＴＲＧ－００２）の組、および（ＳＲＣ－００３，ＴＲＧ－００３）の組の場合は、θ_eに適合する子供の部分木が存在しないため、出力の部分木は構築されない、とする。図１２の（１）は、（ＳＲＣ－００２，ＴＲＧ－００３）の組に対応する部分木、（２）は「（ＳＲＣ－００３，ＴＲＧ－００２）の組に対応する部分木、（３）は（ＳＲＣ－１０３，ＴＲＧ－１０４）の組に対応する部分木である。

【００６２】

次に、図１２の（１）から（３）の出力部分木の単語列リストを展開し、新たな出力単語リストを得る（図１２のステップ３）。そして、各出力単語列に対して、数式６、数式７を用いて翻訳モデル確率、言語モデル確率を算出する。なお、部分木の言語モデル確率は、ここでは、文開始・終了記号を付けずに算出し、図９にない単語列に関しては、例えば、確率「1.0e-7」を割り当てる、とする。

【００６３】

例えば、（ＳＲＣ－００２，ＴＲＧ－００３）の組を用いた場合、「12時に」の出力単語列リストは「at 12 o'clock」，「at noon」，「to noon」であり、「出」の出力単語列リストは「leave」「start」である。したがって、その組み合わせを展開し、翻訳モデル確率、言語モデル確率を算出すると、図１３の出力単語列リストを得る。

図１３において、「入力構文木」は原言語木構造情報の例である。また、「出力構文木」は「目的言語木構造情報」の例である。

同様に、（ＳＲＣ－００３，ＴＲＧ－００２）の組、（ＳＲＣ－１０３，ＴＲＧ－１０４）の組の場合は、図１４の出力単語列リストを得る。

以上の処理により、木構造マッピング確率計算部１７は、図１３、図１４の出力単語列リストを得ることができる。

【００６４】

次に、木構造マッピング確率計算部１７は、入力単語列、入力部分木の構文ラベル、出力部分木の構文ラベルが同一の出力単語列リストが既にバッファに登録されている場合、両者の出力単語列リストをマージする。マージの際、同一の出力単語列が存在する場合は、数式２に従い、翻訳モデル確率の和を算出し、バッファに登録する。

【００６５】

ここで、木構造マッピング確率計算部１７は、例えば、マージの結果、出力単語列リストのサイズが、一定値(Nと表記し、ここでは3を仮定する)を超える場合は、翻訳モデル確率と言語モデル確率の積の上位N個だけを残し、登録する。かかる処理により高速な翻訳処理が可能となる。

上記例の場合、マージの結果、図１５出力単語列リストが得られ、例えば、上位3個だけがバッファに登録される（図１２のステップ４）。

【００６６】

上記処理により、新たに「12時に出」の部分翻訳結果が得られるので、これを再帰的に入力文全体について繰り返すことにより、入力文「バスは12時に出ますか」の部分翻訳結果を得ることができる。なお、最終的に得られた出力単語列リストは、文開始・終了記号込みで言語モデル確率が再計算され、翻訳モデル確率との積が最大の出力単語列を、入力文の翻訳結果として出力する。

以上、本実施の形態によれば、単語と句(複数単語)を区別せず翻訳を行うことができる。

また、本実施の形態によれば、句や単語の順序を階層的に入れ替えることができ、構文的に正しい翻訳文を出力することができる。結果的に翻訳品質が向上する。

【0067】

また、本実施の形態によれば、構文木に複数の候補が得られた時にも、コーパスから自動的に得られた原言語・目的言語木構造モデル、木構造マッピングモデルに基づく確率を基に、最適な出力単語列を構成することができる。したがって、シソーラス等は必要としない。

【0068】

また、本実施の形態の具体例によれば、上記ステップＳ２０６の処理（部分木の数が最少となる出力部分木を選択する処理）の例について説明しなかった。かかる処理は、例えば、以下のような処理である。「すみませんバスは12時に出ますか」が入力されると、部分木の数が最小のものとしては、「すみません」、「バスは12時に出ますか」という２つが得られる。そして、それぞれ「excuse me」、「will the bus leave at 12 o'clock」「will the bus leave at noon」などの出力単語列が得られる。そして、それぞれの部分木から、確率最大の出力単語列を取得し、連結して出力すると、「excuse me will the bus leave at 12 o'clock」となる。

【0069】

さらに、本実施の形態において、出力単語列が多数存在する場合にも、入力単語列、入力・出力の部分木の構文ラベルが同じ出力単語列から、確率が上位の単語列だけを残すことを行えば、適切に候補を削減することができ、翻訳速度が向上する。

なお、本実施の形態の具体例において、言語モデルとして単語ｂｉｇｒａｍモデルを用いたが、単語ｔｒｉｇｒａｍモデル、品詞ｔｒｉｇｒａｍモデル等、他の言語モデルを用いてもよい。

【0070】

さらに、本実施の形態における処理は、ソフトウェアで実現しても良い。そして、このソフトウェアをソフトウェアダウンロード等により配布しても良い。また、このソフトウェアをＣＤ－ＲＯＭなどの記録媒体に記録して流布しても良い。なお、本実施の形態における情報処理装置を実現するソフトウェアは、以下のようなプログラムである。つまり、このプログラムは、コンピュータに、翻訳される元の文章の言語である第一言語の木構造に関する情報であり、複数の単語からなり非終端記号を含まない長単位句を一のノードとする木構造の情報を含む原言語木構造情報と、当該原言語木構造情報に対応する木構造の確率を示す情報である原言語木構造確率を有する原言語木構造レコードを１以上有する原言語木構造モデルを格納しており、翻訳結果の文章の言語である第二言語の木構造に関する情報であり、複数の単語からなり非終端記号を含まない長単位句を一のノードとする木構造の情報を含む目的言語木構造情報と、当該目的言語木構造情報に対応する木構造の確率を示す情報である目的言語木構造確率を有する目的言語木構造レコードを１以上有する目的言語木構造モデルを格納しており、原言語木構造情報と目的言語木構造情報との対応を示す情報であるマッピング情報と、当該マッピング情報が示す対応の確率を示す情報であるマッピング確率を有する木構造マッピングレコードを１以上有する木構造マッピングモデルを格納しており、第二言語の２以上の単語の連続した出現に関する確率の情報である単語出現確率を有する単語出現確率を１以上有する言語モデルを格納しおり、第一言語の文章を受け付ける受付ステップと、前記受付ステップで受け付けた文章を構文解析し、当該文章の一部または全部の木構造に関する情報である木構造情報を、順次得る入力文構

文解析ステップと、前記入力文構文解析ステップで得た木構造情報に対応する１以上の原言語木構造確率を取得し、前記入力文構文解析ステップで得た木構造情報に対応する１以上の目的言語木構造情報と１以上のマッピング確率を取得し、前記１以上の目的言語木構造情報のそれぞれに対応する１以上の目的言語木構造確率を取得し、前記取得した目的言語木構造情報に基づいて、当該目的言語木構造情報を構成する２以上の単語の、１以上の単語出現確率を取得し、前記取得した原言語木構造確率、前記取得したマッピング確率、前記取得した目的言語木構造確率、および前記取得した単語出現確率に基づいて、出力の構文木の評価値を算出する木構造マッピング確率計算ステップと、前記木構造マッピング確率計算ステップで算出した評価値に基づいて、出力する第二言語の一部または全部の構文木を決定し、当該決定した構文木に基づいて第二言語の一部または全部の文章である出力情報を取得する最尤列探索ステップと、前記最尤列探索ステップで取得した１以上の出力情報を有する第二言語の文章を出力する出力ステップを実行させるためのプログラム、である。

【００７１】

また、前記木構造マッピング確率計算ステップは、前記入力文構文解析ステップで得た木構造情報に対応する１以上の原言語木構造確率を取得する原言語木構造確率取得ステップと、前記入力文構文解析ステップで得た木構造情報に対応する１以上の目的言語木構造情報と１以上のマッピング確率を取得するマッピング確率取得ステップと、前記１以上の目的言語木構造情報のそれぞれに対応する１以上の目的言語木構造確率を取得する目的言語木構造確率取得ステップと、前記取得した目的言語木構造情報に基づいて、当該目的言語木構造情報を構成する２以上の単語の、１以上の単語出現確率を前記言語モデル格納部から取得する単語出現確率取得ステップと、前記原言語木構造確率、前記マッピング確率、前記目的言語木構造確率、および前記単語出現確率に基づいて、出力の構文木の評価値を算出する評価値算出ステップを具備しても良い。

前記評価値算出ステップは、前記原言語木構造確率、前記マッピング確率、前記目的言語木構造確率、前記単語出現確率の積を算出し、当該積を出力の構文木の評価値とすることは好適である。

【００７２】

さらに、前記最尤列探索ステップにおいて、前記木構造マッピング確率計算ステップで算出した評価値が最大の構文木を、出力する構文木として決定し、当該決定した構文木に基づいて第二言語の一部または全部の文章である出力情報を取得することは好適である。

【００７３】

また、上記各実施の形態において、各処理（各機能）は、単一の装置（システム）によって集中処理されることによって実現されてもよく、あるいは、複数の装置によって分散処理されることによって実現されてもよい。

【００７４】

また、図１６は、本明細書で述べたプログラムを実行して、上述した種々の実施の形態の情報処理装置を実現するコンピュータの外観を示す。上述の実施の形態は、コンピュータハードウェア及びその上で実行されるコンピュータプログラムで実現され得る。図１６は、このコンピュータシステム３００の概観図であり、図１７は、システム３００のブロック図である。

【００７５】

図１６において、コンピュータシステム３００は、ＦＤ（Ｆｌｅｘｉｂｌｅ　Ｄｉｓｋ）ドライブ、ＣＤ－ＲＯＭ（Ｃｏｍｐａｃｔ　Ｄｉｓｋ　Ｒｅａｄ　Ｏｎｌｙ　Ｍｅｍｏｒｙ）ドライブを含むコンピュータ３０１と、キーボード３０２と、マウス３０３と、モニタ３０４とを含む。

【００７６】

図１７において、コンピュータ３０１は、ＦＤドライブ３０１１、ＣＤ－ＲＯＭドライブ３０１２に加えて、ＣＰＵ（Ｃｅｎｔｒａｌ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ）３０１３と、ＣＰＵ３０１３、ＣＤ－ＲＯＭドライブ３０１２及びＦＤドライブ３０１１に接

続されたバス３０１４と、ブートアッププログラム等のプログラムを記憶するためのＲＯＭ（Ｒｅａｄ－Ｏｎｌｙ　Ｍｅｍｏｒｙ）３０１５と、ＣＰＵ３０１３に接続され、アプリケーションプログラムの命令を一時的に記憶するとともに一時記憶空間を提供するためのＲＡＭ（Ｒａｎｄｏｍ　Ａｃｃｅｓｓ　Ｍｅｍｏｒｙ）３０１６と、アプリケーションプログラム、システムプログラム、及びデータを記憶するためのハードディスク３０１７とを含む。ここでは、図示しないが、コンピュータ３０１は、さらに、ＬＡＮへの接続を提供するネットワークカードを含んでも良い。

【0077】

コンピュータシステム３００に、上述した実施の形態の情報処理装置の機能を実行させるプログラムは、ＣＤ－ＲＯＭ３１０１、またはＦＤ３１０２に記憶されて、ＣＤ－ＲＯＭドライブ３０１２またはＦＤドライブ３０１１に挿入され、さらにハードディスク３０１７に転送されても良い。これに代えて、プログラムは、図示しないネットワークを介してコンピュータ３０１に送信され、ハードディスク３０１７に記憶されても良い。プログラムは実行の際にＲＡＭ３０１６にロードされる。プログラムは、ＣＤ－ＲＯＭ３１０１、ＦＤ３１０２またはネットワークから直接、ロードされても良い。

【0078】

プログラムは、コンピュータ３０１に、上述した実施の形態の情報処理装置の機能を実行させるオペレーティングシステム（ＯＳ）、またはサードパーティープログラム等は、必ずしも含まなくても良い。プログラムは、制御された態様で適切な機能（モジュール）を呼び出し、所望の結果が得られるようにする命令の部分のみを含んでいれば良い。コンピュータシステム３００がどのように動作するかは周知であり、詳細な説明は省略する。

また、上記プログラムを実行するコンピュータは、単数であってもよく、複数であってもよい。すなわち、集中処理を行ってもよく、あるいは分散処理を行ってもよい。

本発明は、以上の実施の形態に限定されることなく、種々の変更が可能であり、それらも本発明の範囲内に包含されるものであることは言うまでもない。

【産業上の利用可能性】

【0079】

以上のように、本発明にかかる機械翻訳装置は、高品質かつ高速な翻訳が可能となる、という効果を有し、機械翻訳装置等として有用である。

【図面の簡単な説明】

【0080】

【図１】実施の形態１における機械翻訳装置のブロック図

【図２】同機械翻訳装置の動作について説明するフローチャート

【図３】同評価処理の動作について説明するフローチャート

【図４】同翻訳モデル確率を算出する処理について説明するフローチャート

【図５】同言語モデル確率を算出する処理について説明するフローチャート

【図６】同原言語木構造モデルの例を示す図

【図７】同目的言語木構造モデルの例を示す図

【図８】同木構造マッピングモデルの例を示す図

【図９】同言語モデルの例を示す図

【図１０】同構文トランスファ方式の機械翻訳を説明する図

【図１１】同機械翻訳装置の初期状態を示す図

【図１２】同機械翻訳装置が構築する部分木を示す図

【図１３】同機械翻訳装置が構成する出力単語列リストを示す図

【図１４】同機械翻訳装置が構成する出力単語列リストを示す図

【図１５】同機械翻訳装置が構成する出力単語列リストを示す図

【図１６】同機械翻訳装置を実現するコンピュータの外観図

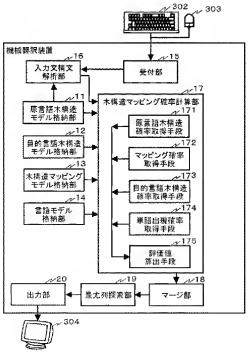
【図１７】同機械翻訳装置のブロック図

【符号の説明】

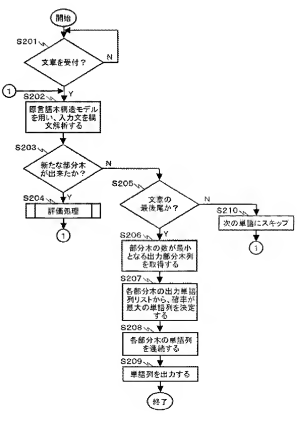
【0081】

１１　原言語木構造モデル格納部
１２　目的言語木構造モデル格納部
１３　木構造マッピングモデル格納部
１４　言語モデル格納部
１５　受付部
１６　入力文構文解析部
１７　木構造マッピング確率計算部
１７１　マッピング確率取得手段
１７２　単語出現確率取得手段
１７３　評価値算出手段

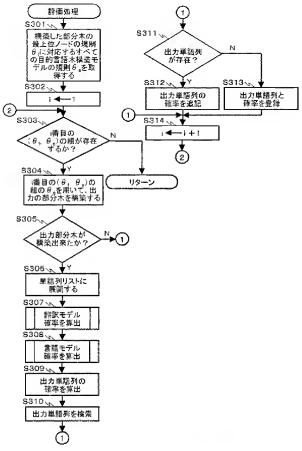
【図1】



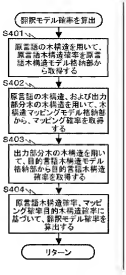
【図2】



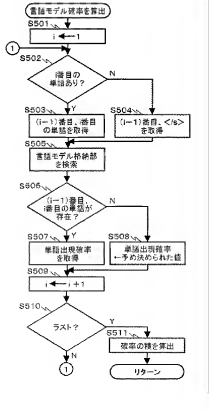
【図３】



【図４】



【図５】



【図６】

名前	原言語木構造情報 親ノード	子ノード列	原言語木構造確率P_{SF}
SRC-001	S	X_{NP}はV$_{VB}$ますか	3.67e-3
SRC-002	VP	X_{PP} Y_V	2.73e-2
SRC-003	VP	X_{PP}にY_V	2.13e-3
SRC-004	NP	X_{NOUN}	9.50e-2
SRC-005	NP	X_{NUM} Y_{NOUN}	1.78e-2
SRC-006	PP	X_{NP} Y_P	8.64e-2
SRC-007	NOUN	バス	1.63e-3
SRC-008	NOUN	列車	6.12e-3
SRC-009	NOUN	時	3.05e-2
SRC-010	NUM	12	1.04e-1
SRC-011	V	出	7.00e-3
SRC-012	P	に	3.85e-1
SRC-101	NP	12時	7.03e-5
SRC-102	PP	12時に	2.30e-5
SRC-103	VP	12時に出	6.59e-6
SRC-104	NP	バス	1.69e-3
SRC-105	S	バスは12時に出ますか	9.16e-6
SRC-106	S	すみません	7.65e-2

【図７】

名前	目的言語木構造情報 親ノード → 子ノード列	目的言語木構造確率 $P(\theta_e)$
TRG-001	SQ → will X_{SQ} Y_{VP}	1.21e-3
TRG-002	VP → Y_{VP} X_{NP}	4.94e-2
TRG-003	VP → Y_{VP} X_{PP}	1.34e-2
TRG-004	NP → the X_{NN}	2.35e-2
TRG-005	NP → a X_{NN}	1.12e-2
TRG-006	NP → X_{NN}	1.37e-2
TRG-007	NP → X_{CD} Y_{NN}	1.95e-3
TRG-008	PP → X_{IN} Y_{NP}	2.32e-1
TRG-009	NN → bus	1.81e-2
TRG-010	NN → train	1.02e-2
TRG-011	NN → o'clock	5.50e-1
TRG-012	CD → 12	9.93e-2
TRG-013	VB → leave	1.67e-2
TRG-014	VB → start	4.37e-3
TRG-015	IN → at	5.82e-2
TRG-016	IN → to	3.51e-1
⋮		
TRG-101	NP → 12 o'clock	1.21e-5
TRG-102	NP → noon	2.32e-4
TRG-103	PP → at 12 o'clock	2.09e-5
TRG-104	VP → leave at 12 o'clock	6.59e-6
TRG-105	NP → the bus	1.26e-3
TRG-106	NP → will the bus leave at 12 o'clock	6.62e-5
TRG-107	S → excuse me	9.46e-3
⋮		

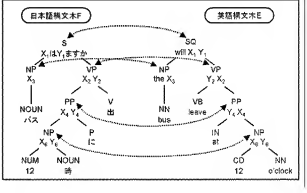
【図８】

マッピング情報 目的言語	原言語	マッピング確率 $P(\theta_{fe})$
SRC-001	TRG-001	4.38e-1
SRC-002	TRG-002	4.24e-2
SRC-002	TRG-003	2.87e-1
SRC-003	TRG-002	5.87e-3
SRC-003	TRG-003	3.44e-3
SRC-004	TRG-004	7.97e-1
SRC-004	TRG-005	6.81e-1
SRC-004	TRG-006	8.80e-1
SRC-005	TRG-007	8.33e-1
SRC-006	TRG-008	6.42e-1
SRC-007	TRG-009	9.83e-1
SRC-008	TRG-010	5.95e-1
SRC-009	TRG-011	9.66e-1
SRC-010	TRG-012	1.0
SRC-011	TRG-013	2.21e-1
SRC-011	TRG-014	5.74e-2
SRC-012	TRG-015	3.75e-1
SRC-012	TRG-016	3.94e-1
⋮		
SRC-101	TRG-101	1.0
SRC-101	TRG-102	1.25e-1
SRC-102	TRG-103	1.0
SRC-103	TRG-104	1.0
SRC-104	TRG-105	6.46e-1
SRC-105	TRG-106	2.50e-1
SRC-106	TRG-107	5.91e-1
⋮		

【図９】

w_e^{i-1}	w_e^i	$P(w_e^i \mid w_e^{i-1})$
<s>	will	2.09e-01
<s>	the	1.74e-01
will	the	2.15e-01
will	a	7.02e-02
the	bus	1.60e-01
a	bus	1.28e-01
bus	leave	2.27e-01
leave	at	2.08e-01
leave	12	2.92e-02
at	12	7.81e-02
at	noon	9.00e-02
12	o'clock	5.25e-01
o'clock	</s>	7.20e-01
noon	</s>	7.31e-01
excuse	me	9.93e-01
⋮	⋮	⋮

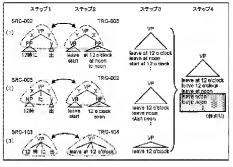
【図１０】



【図１１】

入力構文木	出力構文木	訓訳モデル確率	言語モデル確率
NP → 12時	NP → noon	4.94e-9	1.0
	NP → 12 o'clock	5.79e-9	5.25e-1
PP → 12時に	PP → at 12 o'clock	4.79e-10	4.10e-2
	PP → at noon	5.08e-13	9.00e-2
	PP → to noon	3.22e-12	1.0e-7
V → 出	VB → leave	2.58e-5	1.0
	VB → start	1.76e-6	1.0

【図12】



【図13】

入力構文木	出力構文木	翻訳モデル確率	言語モデル確率
VP → 12時に出る	VP → leave at 12 o'clock	1.30e-18	8.53e-3
	VP → leave at noon	1.36e-21	1.87e-2
	VP → leave to noon	8.73e-21	1.0e-14
	VP → start at 12 o'clock	8.83e-20	4.10e-9
	VP → start at noon	9.36e-23	9.00e-9
	VP → start to noon	5.93e-22	1.0e-14

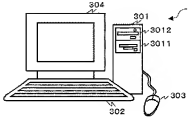
【図14】

入力構文木	出力構文木	翻訳モデル確率	言語モデル確率
VP → 12時に出る	VP → leave 12 o'clock	9.24e-20	1.53e-2
	VP → leave noon	7.69e-20	1.0e-7
	VP → start 12 o'clock	6.27e-21	5.25e-8
	VP → start noon	5.36e-21	1.0e-7
VP → 12時に出る	VP → leave at 12 o'clock	4.34e-11	8.53e-3

【図15】

入力構文木	出力構文木	翻訳モデル確率	言語モデル確率	積
VP → 12時に出る	VP → leave at 12 o'clock	4.34e-11	8.53e-3	3.70e-13
	VP → leave 12 o'clock	9.24e-20	1.53e-2	1.42e-21
	VP → leave at noon	1.36e-21	1.87e-2	2.56e-23
	VP → leave noon	7.69e-20	1.0e-7	7.69e-27
	VP → start noon	5.36e-21	1.0e-7	5.36e-28
	VP → start at 12 o'clock	8.83e-20	4.10e-9	3.62e-28
	VP → start 12 o'clock	6.27e-21	5.25e-8	3.30e-28
	VP → start at noon	9.36e-23	9.00e-9	8.42e-31
	VP → leave to noon	8.73e-21	1.0e-14	8.72e-35
	VP → start to noon	5.93e-22	1.0e-14	5.93e-36

【図16】



【図17】